

**What is claimed is:**

1. A robotics visual and auditory system comprising;
  - a plurality of acoustic models consisting of words spoken by each speaker and their directions combined,
  - a speech recognition engine for executing speech recognition processes to separated sound signals from respective sound sources by using the acoustic models, and
  - a selector for integrating a plurality of speech recognition process results obtained for each of the acoustic models by the speech recognition process, and selecting any one of speech recognition process results,
  - wherein the words spoken simultaneously by each speaker are each recognized.
2. A robotics visual and auditory system as set forth in Claim 1, wherein the selector is made up as to select the speech recognition process results by majority vote.
3. A robotics visual and auditory system as set forth in Claim 1 or Claim 2, wherein it is provided with a dialogue part to output the speech recognition process results selected by the selector to outside.
4. A robotics visual and auditory system comprising;
  - an auditory module which is provided at least with a pair of microphones to collect external sounds, and, based on sound signals from the microphones, determines a direction of at least one speaker by sound source separation and localization by grouping based on pitch extraction and harmonic sounds,
  - a face module which is provided a camera to take images of a robot's front, identifies each speaker, and extracts his face event from each speaker's face recognition and localization, based on images taken by the camera,
  - a motor control module which is provided with a drive motor to rotate the robot in the horizontal direction, and extracts motor event, based on a

rotational position of the drive motor,

an association module which determines each speaker's direction, based on directional information of sound source localization of the auditory event and face localization of the face event, from said auditory, face, and motor events, generates an auditory stream and a face stream by connecting said events in the temporal direction using a Kalman filter for determinations, and further generates an association stream associating these streams, and

an attention control module which conduct an attention control based on said streams, and drive-controls the motor based on an action planning results accompanying the attention control,

wherein the auditory module collects sub-bands having interaural phase difference (IPD) or interaural intensity difference (IID) within a predetermined range by an active direction pass filter having a pass range which, according to auditory characteristics, becomes minimum in the frontal direction, and larger as the angle becomes wider to the left and right, based on an accurate sound source directional information from the association module, and conducts sound source separation by restructuring a wave shape of a sound source, conducts speech recognition of separated sound signals from respective sound sources using a plurality of acoustic models, integrates speech recognition results from each acoustic model by a selector, and judges the most reliable speech recognition result among the speech recognition results.

##### 5. A robotics visual and auditory system comprising;

an auditory module which is provided at least with a pair of microphones to collect external sounds, and, based on sound signals from the microphones, determines a direction of at least one speaker by sound source separation and localization by grouping based on pitch extraction and harmonic sounds,

a face module which is provided a camera to take images of a robot's front, identifies each speaker, and extracts his face event from each speaker's face recognition and localization, based on images taken by the camera,

a stereo module which extracts and localizes a longitudinally long matter, based on a parallax extracted from images taken by a stereo camera, and extracts stereo event,

a motor control module which is provided with a drive motor to rotate the robot in the horizontal direction, and extracts motor event, based on a rotational position of the drive motor,

an association module which determines each speaker's direction, based on directional information of sound source localization of the auditory event and face localization of the face event, from said auditory, face, stereo, and motor events, generates an auditory stream, a face stream and a stereo visual stream by connecting said events in the temporal direction using a Kalman filter for determinations, and further generates an association stream associating these streams, and

an attention control module which conduct an attention control based on said streams, and drive-controls the motor based on an action planning results accompanying the attention control,

wherein the auditory module collects sub-bands having interaural phase difference (IPD) or interaural intensity difference (IID) within a predetermined range by an active direction pass filter having a pass range which, according to auditory characteristics, becomes minimum in the frontal direction, and larger as the angle becomes wider to the left and right, based on an accurate sound source directional information from the association module, and conducts sound source separation by restructuring a wave shape of a sound source, conducts speech recognition of separated sound signals from respective sound sources using a plurality of acoustic models, integrates speech recognition results from each acoustic model by a selector, and judges the most reliable speech recognition result among the speech recognition results.

6. A robotics visual and auditory system as set forth in Claim 4 or Claim 5, characterized in that:

when the speech recognition by the auditory module failed, the attention control module is made up as to collect speeches again from the microphones with the microphones and the camera turned to the sound

source direction of the sound signals, and to perform again speech recognition of the speech by the auditory module, based on the sound signals conducted sound source localization and sound source separation.

7. A robotics visual and auditory system as set forth in Claim 4 or Claim 5, characterized in that:

the auditory module refers to the face event from the face module upon performing the speech recognition.

8. A robotics visual and auditory system as set forth in Claim 5, characterized in that:

the auditory module refers to the stereo event from the stereo module upon performing the speech recognition.

9. A robotics visual and auditory system as set forth in Claim 5, characterized in that:

the auditory module refers to the face event from the face module and the stereo event from the stereo module upon performing the speech recognition.

10. A robotics visual and auditory system as set forth in Claim 4 or Claim 5, wherein it is provided with a dialogue part to output the speech recognition results judged by the auditory module to outside.

11. A robotics visual and auditory system as set forth in Claim 4 or Claim 5, wherein a pass range of the active direction pass filter can be controlled for each frequency.